

Spatial Matérn fields driven by non-Gaussian noise

David Bolin

*Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Sweden
e-mail: bolin@maths.lth.se*

Abstract: The article studies non-Gaussian extensions of a recently discovered link between certain Gaussian random fields, expressed as solutions to stochastic partial differential equations (SPDEs), and Gaussian Markov random fields. The focus is on non-Gaussian random fields with Matérn covariance functions, and in particular we show how the SPDE formulation of a Laplace moving average model can be used to obtain an efficient simulation method as well as an accurate parameter estimation technique for the model. This should be seen as a demonstration of how these techniques can be used, and generalizations to more general SPDEs are readily available.

AMS 2000 subject classifications: 62M40, 62H11, 60H15.

Keywords and phrases: Matérn covariances, SPDE, Laplace moving averages, Markov random fields, process convolutions, EM algorithm.

1. Introduction

Recently, [Lindgren, Rue and Lindström \(2011\)](#) derived a link between certain Gaussian fields, that can be represented as solutions to stochastic partial differential equations (SPDEs), and Gaussian Markov random fields (GMRFs). The main idea is to approximate these Gaussian fields using basis expansions $\sum_i w_i \varphi_i(s)$ where the stochastic weights $\{w_i\}$ are calculated using the stochastic weak formulation of the corresponding SPDE. For certain choices of the basis functions $\{\varphi_i\}$, especially compactly supported functions, the weights form GMRFs. Because of the Markov property of the weights, fast numerical techniques for sparse matrices can be used when estimating parameters and doing spatial prediction in these models. This greatly improves the applicability to problems involving large data sets, where traditional methods in statistics fail due to computational issues. However, the advantages of representing Gaussian fields as solutions to SPDEs are not only computational. Using the SPDE representation, non-stationary extensions are easily obtained by allowing spatially varying parameters in the SPDE ([Lindgren, Rue and Lindström, 2011](#)), and the model class can be generalized to include more general covariance structures by generalizing the class of generating SPDEs ([Bolin and Lindgren, 2011](#)). These are indeed useful features from an applied point of view as many applications require complicated non-stationary models to accurately capture the covariance structure of the data.

So far these methods have only been used in Gaussian settings, and it has not been clear whether they are applicable when the Gaussianity assumption cannot be justified. Therefore, this work will focus on extending the SPDE methods beyond Gaussianity. A new type of non-Gaussian models that has proved to be useful in practical applications is the Laplace moving average models (Åberg, Podgórski and Rychlik, 2009, Åberg and Podgórski, 2011). These are processes obtained by convolving some deterministic kernel function with stochastic Laplace noise. The models share many good properties with the Gaussian models while allowing for heavier tails and asymmetry in the data, making them interesting alternatives in practical applications (see e.g. Bogsjö, Podgórski and Rychlik, 2012). One of the motivating examples in Åberg and Podgórski (2011) is a Laplace moving average model with Matérn covariances. This model can be seen as the solution to the same SPDE that generates Gaussian Matérn field but where the Gaussian white noise forcing is replaced with Laplace noise. It has previously been shown that the SPDE model formulation of Gaussian Matérn fields has many computational advantages compared with the process convolution formulation (Bolin and Lindgren, 2009, Simpson, Lindgren and Rue, 2010). We demonstrate here that for the Laplace moving average models, the SPDE formulation can also be used to derive a new likelihood-based parameter estimation technique as well as an efficient simulation procedure.

The structure of the paper is as follows. Section 2 contains an introduction to the Matérn covariance family and the SPDE formulation in the Gaussian case. In Section 3, stochastic Laplace fields are introduced, and some properties of the Laplace-driven SPDE model are derived. Subsequently, in Section 4, the Markov approximation technique by Lindgren, Rue and Lindström (2011) is extended to the Laplace model, and its sampling is discussed in Section 5. A parameter estimation technique based on the EM algorithm is derived in Section 6, and Section 7 contains a simulation study showing that it gives reliable parameter estimates. Finally, Section 8 contains a summary and discussion of future work and possible extensions.

2. Gaussian Matérn fields

The Matérn covariance family (Matérn, 1960) is often used when modeling spatial data. There are a few different parameterizations of the Matérn covariance function in the literature, and the one most suitable in our context is

$$C(\mathbf{h}) = \frac{2^{1-\nu}\phi^2}{(4\pi)^{\frac{d}{2}}\Gamma(\nu + \frac{d}{2})\kappa^{2\nu}}(\kappa\|\mathbf{h}\|)^\nu K_\nu(\kappa\|\mathbf{h}\|), \quad \mathbf{h} \in \mathbb{R}^d, \quad (1)$$

where d is the dimension of the domain, ν is a shape parameter, κ^2 a scale parameter, ϕ^2 a variance parameter, and K_ν is a modified Bessel function of the second kind of order $\nu > 0$. The associated spectrum is

$$S(\mathbf{k}) = \frac{\phi^2}{(2\pi)^d} \frac{1}{(\kappa^2 + \mathbf{k}^\top \mathbf{k})^{\nu + \frac{d}{2}}}. \quad (2)$$

As the properties of Gaussian fields are given by their first two moments, the standard way of specifying Gaussian Matérn fields is to choose the mean value, $\mu(\mathbf{s})$ possibly spatially varying, and then let the covariance function be of the form (1). An alternative way of specifying a Gaussian field on \mathbb{R}^d is to view it as a process convolution

$$X(\mathbf{s}) = \int_{\mathbb{R}^d} k(\mathbf{s}, \mathbf{u}) \mathcal{B}(\mathrm{d}\mathbf{u}), \quad (3)$$

where k is some deterministic kernel function and \mathcal{B} is a Brownian sheet (Higdon, 2001). One of the advantages with this construction is that non-stationary extensions are easily constructed by allowing the convolution kernel to be dependent on the location \mathbf{s} . If, however, the process is stationary, the kernel k depends only on $\mathbf{s} - \mathbf{u}$ and the covariance function for X is

$$C(\mathbf{h}) = \int_{\mathbb{R}^d} k(\mathbf{u} - \mathbf{h}) k(\mathbf{u}) \mathrm{d}\mathbf{u}.$$

Thus, the covariance function C , the spectrum S , and the kernel k are related through

$$(2\pi)^d |\mathcal{F}(k)|^2 = \mathcal{F}(C) = S,$$

where $\mathcal{F}(\cdot)$ denotes the Fourier transform. Since the spectral density for a Matérn field in dimension d with parameters ν , ϕ^2 , and κ is given by (2), one finds that the corresponding symmetric non-negative kernel is a Matérn covariance function with parameters $\nu_k = \frac{\nu}{2} - \frac{d}{4}$, $\phi_k = \sqrt{\phi}$, and $\kappa_k = \kappa$.

In yet another setting, Gaussian Matérn fields can be viewed as the solution to the SPDE

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} X(\mathbf{s}) = \phi \mathcal{W}(\mathbf{s}), \quad (4)$$

where $\mathcal{W}(\mathbf{s})$ is Gaussian white noise, $\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial s_i^2}$ is the Laplace operator, and $\alpha = \nu + d/2$ (Whittle, 1963). As discussed in Lindgren, Rue and Lindström (2011), there is an implicit assumption of appropriate boundary conditions needed if one wants the solutions to be stationary Matérn fields.

The connection between (3) and (4) is through the Green's function of the differential operator in (4)

$$G_\alpha(\mathbf{s}, \mathbf{t}) = \frac{2^{1-\frac{\alpha-d}{2}}}{(4\pi)^{\frac{d}{2}} \Gamma(\frac{\alpha}{2}) \kappa^{\alpha-d}} (\kappa \|\mathbf{s} - \mathbf{t}\|)^{\frac{\alpha-d}{2}} K_{\frac{\alpha-d}{2}}(\kappa \|\mathbf{s} - \mathbf{t}\|), \quad (5)$$

that serves as a kernel in (3). It is straightforward to show that $G_\alpha \in L_p(\mathbb{R}^d)$ if and only if $\alpha > \frac{(p-1)d}{p}$ (see for example Samko, Kilbas and Marichev (1992) p. 538), and in particular $\alpha > d/2$ guarantees that $G_\alpha \in L_2(\mathbb{R}^d)$.

A non-Gaussian model with Matérn covariances could be constructed either using the process convolution formulation (3) where the Brownian sheet is replaced by some non-Gaussian process, or through the SPDE formulation (4) with non-Gaussian noise. Such non-Gaussian extensions are discussed next.

3. Non-Gaussian SPDE-based models

A simple way of moving beyond Gaussianity in the SPDE model (4) is to allow for a stochastic variance parameter ϕ . By choosing ϕ as an inverse-gamma distributed random variable, the resulting field has t-distributed marginal distributions and is therefore sometimes referred to as a t-distributed random field (Røislien and Omre, 2006). In a Bayesian setting, this extension can be interpreted simply as choosing a certain prior distribution for the variance, and one can of course come up with many other non-Gaussian models by changing this distribution. However, models constructed in this way are non-Gaussian only in a very limited sense. Namely, every realization of them behaves exactly as a Gaussian field with a globally re-scaled variance, and because of this, they are all non-ergodic as the parameters in the prior distribution cannot be estimated from a single realization of the field. One would prefer a non-Gaussian model where the actual sample paths behave differently from a stationary Gaussian field, and one way of achieving this is to let the variance parameter be spatially and stochastically varying. Both Lindgren, Rue and Lindström (2011) and Bolin and Lindgren (2011) explore this option by expressing $\log \phi(\mathbf{s})$ as a regression on a few known basis functions where the stochastic weights are estimated from data. This was interpreted as a non-stationary Gaussian model, but could also be viewed as a, somewhat limited, non-Gaussian model with a slowly spatially varying variance parameter $\phi(\mathbf{s})$. To obtain a model which is intrinsically non-Gaussian also within realizations, one can draw $\phi(\mathbf{s})$ at random independently for each \mathbf{s} . The right-hand side of (4) is then a product of two independent noise fields. The following non-Gaussian models essentially can be interpreted as a formal realization of this idea.

One interesting type of distributions, obtained by taking a random variance and mean in an otherwise Gaussian random variable, are the generalized asymmetric Laplace distributions (Åberg, Podgórski and Rychlik, 2009). The Laplace distribution is defined through the characteristic function with parameters $\mu, \gamma \in \mathbb{R}$ and $\sigma, \tau > 0$

$$\varphi(u) = e^{i\gamma u} \left(1 - i\mu u + \frac{\sigma^2}{2} u^2 \right)^{-\tau}.$$

The distribution is symmetric if $\mu = 0$ and asymmetric otherwise. The shape of the distribution is governed by τ and the scale by σ . The distribution is infinitely divisible, and a useful characterization is that if Z is a standard normal variable and Γ is an independent gamma variable with shape τ , then $\gamma + \mu\Gamma + \sigma\sqrt{\Gamma}Z$ has an asymmetric Laplace distribution.

Stochastic Laplace noise can now be obtained from an independently scattered random measure Λ , defined for a Borel set B in \mathbb{R}^d by the characteristic function

$$\varphi_{\Lambda(B)}(u) = e^{i\gamma m(B)u} \left(1 - i\mu u + \frac{\sigma^2}{2} u^2 \right)^{-m(B)},$$

where the measure m is referred to as the control measure of Λ . This does not define Laplace noise in a direct manner, but similarly to how Gaussian white

noise can be seen as a differentiated Brownian sheet (Walsh, 1986), Laplace noise can be viewed in the sense of distributions (generalized functions) as a differentiated Laplace field. The most transparent characterization is through the following series representation of the Laplace field $\Lambda(\mathbf{s})$ on a compact set $D \in \mathbb{R}^d$:

$$\Lambda(\mathbf{s}) = \gamma \mathbf{s} + \sum_{k=1}^{\infty} \left(\Gamma_k + G_k \sqrt{\Gamma_k} \right) \mathbf{1}(\mathbf{s} \geq \mathbf{s}_k), \quad \mathbf{s} \in D, \quad (6)$$

where G_k are iid $N(0, 1)$ random variables, \mathbf{s}_k are iid uniform random variables on D , and

$$\mathbf{1}(\mathbf{s} \geq \mathbf{s}_k) = \begin{cases} 1 & \text{if } s_i \geq s_{k,i} \text{ for all } i \leq d, \\ 0 & \text{otherwise.} \end{cases}$$

The random variables Γ_k can be written as $\Gamma_k = e^{-\nu \gamma_k} W_k$ where W_k are iid standard exponential variables and γ_k are the arrival times of a Poisson process with intensity 1. Thus, Laplace noise can be expressed as a distribution (generalized function)

$$\dot{\Lambda} = \gamma + \sum_{k=1}^{\infty} \left(\Gamma_k + G_k \sqrt{\Gamma_k} \right) \delta_{\mathbf{s}_k}, \quad (7)$$

where $\delta_{\mathbf{s}_k}$ is the Dirac delta distribution centered at \mathbf{s}_k .

The model of interest is the solution X to the Laplace-driven SPDE

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} X = \dot{\Lambda}, \quad (8)$$

where both X and $\dot{\Lambda}$ are viewed as random variables valued in the space of tempered distributions. To clarify in what way the solution to this equation exists, we look at a general SPDE

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} X = \dot{M}, \quad (9)$$

where M is an arbitrary independently scattered L_2 -valued random measure with $\mathbb{E}(|M(\mathrm{d}\mathbf{x})|^2) = C \mathrm{d}\mathbf{x}$ for some constant $C < \infty$. Examples of such measures are the Laplace measures of interest here but also standard Brownian sheets. As usual for fractional Laplacian operators (Samko, Kilbas and Maricev, 1992), $\mathcal{T} = (\kappa^2 - \Delta)^{\frac{\alpha}{2}}$ is defined using the Fourier transform through $\mathcal{F}(\mathcal{T}f) = \mathcal{P}\hat{f}$, where \hat{f} is the Fourier transform of the function f , $(\mathcal{P}\hat{f})(\mathbf{k}) = (\kappa^2 + \mathbf{k}^\top \mathbf{k})^{\frac{\alpha}{2}} \hat{f}(\mathbf{k})$, and the operator \mathcal{T} is well-defined for example for all $f \in L_p(\mathbb{R}^d)$ for $1 \leq p \leq \infty$. The definition applies also when f is a distribution or, more specifically, a tempered distribution. Thus, (9) is viewed as an equation for two random (tempered) distributions so the equation has to be interpreted in the weak sense

$$\mathcal{T}X(\varphi) = \dot{M}(\varphi), \quad (10)$$

where φ is in some appropriate space of test functions. Now, the action of the self-adjoint operator \mathcal{T} can be moved to the test function on the left-hand side

and (10) can be rewritten in a more explicit fashion as

$$X(\mathcal{T}\varphi, \omega) = \int \varphi(\mathbf{s}) M(\mathrm{d}\mathbf{s}, \omega). \quad (11)$$

Here, we have included the second argument $\omega \in \Omega$ to highlight that the sought functional X is random, and the equation should hold for ω in a certain full probability set $\Omega_0 \in \Omega$ and universally for each φ .

To describe the solutions of (9), we need the Sobolev spaces H_n of fractional order n . These are usually defined using the Fourier transform in the following way. Let E be the Schwartz space of rapidly decreasing functions on \mathbb{R}^d , for $u \in E'$ (the dual of E , also referred to as the space of tempered distributions), define the Fourier transform of u as $\hat{u}(\varphi) = u(\hat{\varphi})$, where $\hat{\varphi}$ is the usual Fourier transform on \mathbb{R}^d of $\varphi \in E$. Define a norm on E by

$$\|u\|_n = \int_{\mathbb{R}^d} (1 + |\mathbf{k}|^2)^n |\hat{u}(\mathbf{k})|^2 \mathrm{d}\mathbf{k}$$

and let H_n be the completion of E in this norm. By Plancherel's theorem, one has that $H_0 = L_2(\mathbb{R}^d)$ and one can show that for the special case $n \in \mathbb{N}$, H_n is identical to the classical Sobolev space of L_2 functions with all partial derivatives of order n or less in L_2 . The space H_{-n} is the dual space of H_n and does in general contain distributions.

Let us note that the right hand side of (11) in principle may not be defined on a full probability set uniformly for all φ . However, one can regularize M so that $\varphi \rightarrow M(\varphi)$ is in fact a random distribution. Indeed, since

$$\mathbb{E}(|M(\varphi)|^2) = C \int \varphi(\mathbf{s})^2 \mathrm{d}\mathbf{s} = C \|\varphi\|_0^2,$$

the random linear functional $\varphi \rightarrow M(\varphi)$ is continuous in probability on H_n for any $n \geq 0$, and by Theorem 4.1 in Walsh (1986) there exists a version of M which is almost surely in H_{-n} for $n > d/2$. From now on we always assume that we deal with such a version.

Following Walsh (1986), we say that $X(\cdot, \omega)$ is an H_n -solution of (9) if for a.e. ω , $X(\cdot, \omega)$ is an element of H_{-n} and (11) holds for every $\varphi \in H_n$. In other words, we aim at finding a random functional X that almost surely is a distribution and satisfies (9) as a continuous functional on H_n . The proof of the following proposition is similar to the proof of Proposition 9.1 in Walsh (1986) where the existence of the solution to the stochastic Poisson equation on a bounded domain in \mathbb{R}^d was demonstrated.

Proposition 3.1. *Assume that M is an independently scattered L_2 -valued random measure with $\mathbb{E}(|M(\mathrm{d}\mathbf{x})|^2) = C \mathrm{d}\mathbf{x}$. Then for $\kappa > 0$, $\alpha > 0$, there exists a random functional $X : H_n \times \Omega \rightarrow \mathbb{R}$ such that for a certain set Ω_0 , $P(\Omega_0) = 1$ and for all $\omega \in \Omega_0$ and all $\varphi \in H_n$*

$$X(\varphi, \omega) = \int G^\alpha \varphi(\mathbf{x}) M(\mathrm{d}\mathbf{x}, \omega), \quad (12)$$

where $G^\alpha \varphi(\mathbf{x}) = \int G_\alpha(\mathbf{s}, \mathbf{x}) \varphi(\mathbf{s}) d\mathbf{s}$ and G_α is given by (5). This is the unique H_n -solution to (9) if $n > d/2$, and moreover we have $X \in H_m$ almost surely for $m < \alpha - d/2$.

Proof. From the standard theory of fractional differential equations, one has that G^α maps H_n isomorphically onto $H_{n+\alpha}$ (see e.g. Samko, Kilbas and Marichev, 1992, p.547). Let X be any H_n -solution to (9) and let $\psi = G^\alpha \varphi$. Applying (11) to ψ and using that $\mathcal{T}G^\alpha \varphi = \varphi$ one gets that

$$X(\varphi) = X(\mathcal{T}G^\alpha \varphi) = X(\mathcal{T}\psi) = \int \psi(\mathbf{y}) M(d\mathbf{y}) = \int G^\alpha \varphi(\mathbf{y}) M(d\mathbf{y}).$$

Thus this solution also satisfies (12) and the solution is unique if it exists.

To prove existence, let X be defined by (12) and take $\varphi \in L_2(\mathbb{R}^d)$. Then

$$\begin{aligned} \mathbb{E}(|X(\varphi)|^2) &= \mathbb{E} \left[\left(\int G^\alpha \varphi(\mathbf{y}) M(d\mathbf{y}) \right)^2 \right] \\ &= \int G^\alpha \varphi(\mathbf{x}) G^\alpha \varphi(\mathbf{y}) \mathbb{E}[M(d\mathbf{x}) M(d\mathbf{y})] \\ &= C \int (G^\alpha \varphi(\mathbf{x}))^2 d\mathbf{x} = C \|G^\alpha \varphi\|_0^2 \leq C_2 \|\varphi\|_{-\alpha}^2, \end{aligned}$$

where the last inequality follows from that G^α maps $H_{n-\alpha} \rightarrow H_n$ boundedly for $\alpha > 0$. Thus, it follows that X is a random linear functional that is continuous in probability on $H_{-\alpha}$. The embedding maps $H_{n_1} \rightarrow H_{n_2}$ are of Hilbert-Schmidt type if $n_1 > n_2 + d/2$ (see e.g. Example 1a in Walsh, 1986), and using this with $n_2 = -\alpha$ together with Theorem 4.1 in Walsh (1986) one gets that there exists a version of X which is almost surely in H_{-n} if $n > d/2 > d/2 - \alpha$. From now on, X is such a version and we note that $X \in H_m$ almost surely for $m < \alpha - d/2$.

What is left to show now is that X with probability one satisfies (11) for each $\varphi \in H_n$ for $n > d/2$. To that end, first note that if $\varphi \in H_n$, then by the definitions of \mathcal{T} and G^α one has

$$\begin{aligned} G^\alpha \mathcal{T}\varphi(\mathbf{s}) &= \int G_\alpha(\mathbf{y}, \mathbf{s}) \mathcal{T}\varphi(\mathbf{y}) d\mathbf{y} \\ &= \mathcal{F}^{-1} \left((\kappa^2 + \mathbf{k}^\top \mathbf{k})^{-\frac{\alpha}{2}} (\kappa^2 + \mathbf{k}^\top \mathbf{k})^{\frac{\alpha}{2}} \hat{\varphi}(\mathbf{k}) \right) (\mathbf{s}) \\ &= \varphi(\mathbf{s}). \end{aligned}$$

Let $n > d/2$ and fix $\varphi \in H_n$. If $M(\varphi, \omega)$ denotes the functional $\int \varphi(\mathbf{s}) M(d\mathbf{s}, \omega)$, one has by the definition of X and by the equation above that

$$\int |X(\mathcal{T}\varphi, \omega) - M(\varphi, \omega)|^2 d\mathbb{P}(\omega) = 0.$$

Hence, there is a set $\Omega_\varphi \subset \Omega$ with $\mathbb{P}(\Omega_\varphi) = 1$ such that for each $\omega \in \Omega_\varphi$ one has $X(\mathcal{T}\varphi, \omega) = M(\varphi, \omega)$. Now, H_n is separable, so we can choose a countable

base $B = \{b_i\}_{i=1}^\infty$ in H_n and define $\bar{\Omega}_0 = \cap_{i=1}^\infty \Omega_{b_i}$. Then equality holds for each $f \in B$ and for each $\omega \in \bar{\Omega}_0$ and $P(\bar{\Omega}_0) = 1$ by the countability of B .

The map $\varphi \rightarrow \mathcal{T}\varphi \rightarrow X(\mathcal{T}\varphi)$ of $H_n \rightarrow H_{n-\alpha} \rightarrow \mathbb{R}$ is continuous since X is continuous on H_n for $n > d/2 - \alpha$ and \mathcal{T} is a continuous map from H_n to $H_{n-\alpha}$. Thus, both $X(\mathcal{T}\cdot, \omega)$ and $M(\cdot, \omega)$ are continuous functionals on H_n for ω in some full probability set $\bar{\Omega}_0$ and equality therefore holds in (11) for each $\varphi \in H_n$ for $\omega \in \Omega_0 = \bar{\Omega}_0 \cap \tilde{\Omega}_0$ since B is linearly dense in H_n . \square

Remark 1. By similar arguments one can show that the solution X defined in Proposition 3.1 also is a solution to the SPDE (9) in the sense that with probability one $X \in E'$ and (11) holds for every $\varphi \in E$. This is, however, a weaker statement since $E = \cap_n H_n$ and $E' = \cup_n H_n$.

Remark 2. The solution X defined in Proposition 3.1 is in general a random linear functional. However, it can be identified with a random function if $\alpha > d/2$ since $X \in H_m$ almost surely for $m < \alpha - d/2$. Using the relation between α and the parameter ν in the Matérn covariance function, $\alpha = \nu + d/2$, we see that $X \in H_m$ almost surely for $m < \nu$. Thus, ν acts as a smoothness parameter for the solution since the sample paths almost surely will be differentiable if $\nu > 1$, two times differentiable if $\nu > 2$ etc.

Remark 3. The previous remark can be strengthened using the Sobolev embedding theorem which shows that H_n can be embedded in the Hölder space $C_k^r(\mathbb{R}^d)$ where $n - (r + k) = d/2$ and $r \in (0, 1)$ (see e.g. Adams, 1975). The space $C_k^r(\mathbb{R}^d)$ consists of functions such that all partial derivatives up to order k are continuous and such that the k th partial derivatives are Hölder continuous with exponent r . Thus, if $\nu > d/2$, we almost surely have $X \in C_k^r(\mathbb{R}^d)$ (after possibly redefining it on a set of measure zero) where k is the integer part of $\nu - d/2$ and $r = \nu - d/2 - k$.

We now go back to the special case of Laplace noise and since the main interest here is ordinary random fields with Matérn covariance functions, we from now on assume that $\alpha > d/2$ in (8). One sometimes uses $m(A) = l(A)\tau$, where l is the Lebesgue measure and τ some constant, as a control measure for Λ . By the definition of the differential operator \mathcal{T} , it is then easy to see that the spectrum for the solution X is

$$R_X(\mathbf{k}) = \frac{\tau(\sigma^2 + \mu^2)}{(2\pi)^d} \frac{1}{(\kappa^2 + \mathbf{k}^\top \mathbf{k})^\alpha}.$$

Thus, the covariance function for X is a Matérn covariance of the form (1) with $\phi^2 = \tau(\sigma^2 + \mu^2)$. Since X is Laplace noise convolved with a Green function, which also has the form of a Matérn covariance function, the model is equivalent to the Laplace moving average models in Åberg, Podgórski and Rychlik (2009), Åberg and Podgórski (2011). Thus, using Theorem 1 in Åberg and Podgórski (2011), the marginal distribution for $X(\mathbf{s})$ is given by the characteristic function

$$\phi_X(u) = \exp \left(\tau \int i\gamma G_\alpha(\mathbf{s}, \mathbf{t}) u - \log \left(1 - i\mu u G_\alpha(\mathbf{s}, \mathbf{t}) + \frac{\sigma^2 u^2}{2} G_\alpha^2(\mathbf{s}, \mathbf{t}) \right) d\mathbf{t} \right). \quad (13)$$

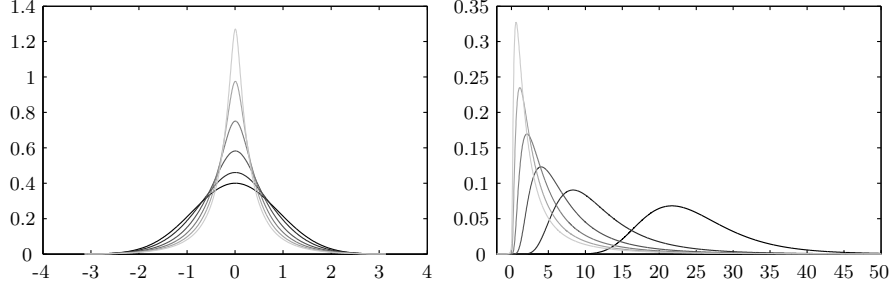


FIGURE 1. Marginal distributions of the solution $X(\mathbf{s})$ to (8) in the case of symmetric (left panel) and asymmetric (right panel) Laplace noise.

A few examples of the marginal distributions for symmetric and asymmetric cases are shown in Figure 1.

4. Hilbert space approximations

To obtain a computationally efficient representation of a Matérn field, the Hilbert space approximation technique by Lindgren, Rue and Lindström (2011) can be used. The starting point is to consider the stochastic weak formulation (10) of the SPDE. A finite element approximation of the solution X is then obtained by representing it as a finite basis expansion $\tilde{X} = \sum_{i=1}^n w_i \varphi_i(\mathbf{s})$, where the stochastic weights are calculated by requiring (10) to hold for only a specific set of test functions $\{\psi_i, i = 1, \dots, n\}$ and $\{\varphi_i\}$ is a set of predetermined basis functions. To simplify the presentation, we first look at the case $\alpha/2 \in \mathbb{N}$ and then turn to the case of a general $\alpha > d/2$.

4.1. The case $\alpha/2 \in \mathbb{N}$

To construct the approximation for $\alpha = 2, 4, \dots$, we first look at the fundamental case $\alpha = 2$. Lindgren, Rue and Lindström (2011) then use $\psi_i = \varphi_i$, and one then has

$$(\kappa^2 - \Delta)\tilde{X}(\varphi_i) = \sum_{j=1}^n w_j \langle \varphi_i, (\kappa^2 - \Delta)\varphi_j \rangle,$$

where $\langle f, g \rangle = \int f(s)g(s) ds$. By introducing the vector $\mathbf{w} = (w_1, \dots, w_n)^\top$ and a matrix \mathbf{K} with elements $\mathbf{K}_{ij} = \langle \varphi_i, (\kappa^2 - \Delta)\varphi_j \rangle$, the left hand side of (10) can be written as $\mathbf{K}\mathbf{w}$. Under mild conditions on the basis functions, one has

$$\begin{aligned} \langle \varphi_i, (\kappa^2 - \Delta)\varphi_j \rangle &= \kappa^2 \langle \varphi_i, \varphi_j \rangle - \langle \varphi_i, \Delta\varphi_j \rangle \\ &= \kappa^2 \langle \varphi_i, \varphi_j \rangle + \langle \nabla\varphi_i, \nabla\varphi_j \rangle. \end{aligned}$$

Hence, the matrix \mathbf{K} can be written as the sum $\mathbf{K} = \kappa^2 \mathbf{C} + \mathbf{G}$ where \mathbf{C} and \mathbf{G} are matrices with elements $\mathbf{C}_{ij} = \langle \varphi_i, \varphi_j \rangle$ and $\mathbf{G}_{ij} = \langle \nabla\varphi_i, \nabla\varphi_j \rangle$ respectively.

4.1.1. Gaussian noise

In the Gaussian case, when \dot{M} is Gaussian white noise, the right hand side of (10) under the finite element approximation can be shown to be Gaussian with mean zero and covariance \mathbf{C} . Thus, one has

$$\mathbf{w} \sim \mathcal{N}(0, \mathbf{K}^{-1}\mathbf{C}\mathbf{K}^{-1}). \quad (14)$$

For higher order $\alpha/2 \in \mathbb{N}$, the weak solution is obtained recursively. If, for example, $\alpha = 4$ the solution to $(\kappa^2 - \Delta)^2 X_0 = \mathcal{W}$ is obtained by solving $(\kappa^2 - \Delta)X_0 = \tilde{X}$, where \tilde{X} is the solution for the case $\alpha = 2$. This results in replacing the matrix \mathbf{K} with a matrix \mathbf{K}_α defined recursively as $\mathbf{K}_\alpha = \mathbf{K}\mathbf{C}^{-1}\mathbf{K}_{\alpha-2}$, where $\mathbf{K}_2 = \mathbf{K}$. For more details about these representations in the Gaussian case, see Lindgren, Rue and Lindström (2011).

So far, we have not specified how the basis functions $\{\varphi_i\}$ should be chosen, but this choice will determine the quality of the approximation as well as some computational properties. If, for example, Daubechies wavelets are used as basis functions, the precision matrix (inverse covariance matrix) \mathbf{Q} for the weights is a sparse matrix (Bolin and Lindgren, 2009), which facilitates the use of efficient sparse matrix techniques when using this model. Lindgren, Rue and Lindström (2011) used piecewise linear basis functions induced by triangulating the domain, and in this case \mathbf{C} is a sparse matrix, but its inverse is dense. To obtain a sparse precision matrix in this case (which is needed for efficient GMRF computations), one can approximate \mathbf{C} with a diagonal matrix $\tilde{\mathbf{C}}$ with elements $\tilde{C}_{ii} = \int \varphi_i(\mathbf{s}) d\mathbf{s}$. To simplify the notation later, we denote the i th element on the diagonal by \mathbf{a}_i as it is the area where $\varphi_i > \varphi_j$ for $j \neq i$. For more details on this approximation and the choice of basis functions, see Bolin and Lindgren (2009).

4.1.2. Laplace noise

For the Laplace case, one has $\dot{M} = \dot{\Lambda}$ in the weak formulation (10). Under the finite element approximation, the left-hand side can, as in the Gaussian case, be written as $\mathbf{K}_\alpha \mathbf{w}$. Using Theorem 1 in Åberg and Podgórski (2011), the distribution of the right-hand side in the case of Laplace noise is given by the characteristic function

$$\phi_\Lambda(\mathbf{u}) = \exp \left(\tau \int_\Omega i\gamma \boldsymbol{\varphi}(\mathbf{s})^\top \mathbf{u} - \log \left(1 - i\mu \boldsymbol{\varphi}(\mathbf{s})^\top \mathbf{u} + \frac{\sigma^2}{2} (\boldsymbol{\varphi}(\mathbf{s})^\top \mathbf{u})^2 \right) d\mathbf{s} \right),$$

where $\boldsymbol{\varphi}(\mathbf{s}) = (\varphi_1(\mathbf{s}), \dots, \varphi_n(\mathbf{s}))^\top$. This representation is not very convenient for approximation and simulation of the model. Instead we will use a representation based on the series expansion (6) of Λ . However, for a moment, we turn to the more general setup of type-G processes to hint at how this technique could be applied also for this broader class of random fields.

Recall that a Lévy process is type G if its increments can be represented as a Gaussian variance mixture $V^{1/2}Z$ where Z is a standard Gaussian variable and

V is a non-negative infinitely divisible random variable. Clearly, the Laplace fields are of type G as their increments are of the form $\Gamma^{1/2}Z$ where Γ is a gamma variable. [Rosiński \(1991\)](#) showed that every Lévy process of type G can be represented as a series expansion similar to the expansion (6) for the Laplace fields. This expansion also holds in \mathbb{R}^d , and for a compact domain $D \in \mathbb{R}^d$ it can be written as

$$M(\mathbf{s}) = \sum_{k=1}^{\infty} G_k g(\gamma_k)^{\frac{1}{2}} \mathbf{1}(\mathbf{s} \geq \mathbf{s}_k),$$

where the function g is the generalized inverse of the tail Lévy measure for V and the other variables are the same as in the Laplace case (6). Since V is infinitely divisible, there exists a non-decreasing Lévy process $V(\mathbf{s})$ with increments distributed the same as V . This process has the series representation

$$V(\mathbf{s}) = \sum_{k=1}^{\infty} g(\gamma_k)^{\frac{1}{2}} \mathbf{1}(\mathbf{s} \geq \mathbf{s}_k). \quad (15)$$

Now, consider the integral of some basis function φ_i with respect to M , which can be represented as

$$\int_D \varphi_i(\mathbf{s}) M(d\mathbf{s}) \stackrel{d}{=} \sum_{k=1}^{\infty} \varphi_i(\mathbf{s}_k) G_k \sqrt{g(\gamma_k)}. \quad (16)$$

Thus, the distribution of $(\int_D \varphi_1(\mathbf{s}) M(d\mathbf{s}), \dots, \int_D \varphi_n(\mathbf{s}) M(d\mathbf{s}))$ can be approximated in distribution by taking partial sums of the series in (16). Another way of calculating the distribution is to evaluate the integrals by conditioning on the variance process $V(\mathbf{s})$ ([Wiktorsson, 2002](#)); given that $\int_D \varphi_i^2(\mathbf{s}) V(d\mathbf{s}) < \infty$, the integral conditionally on V is simply a Gaussian variable

$$\int_D \varphi_i(\mathbf{s}) M(d\mathbf{s}) | V \sim \mathbf{N} \left(0, \int_D \varphi_i^2(\mathbf{s}) V(d\mathbf{s}) \right).$$

Going back to the case of Laplace noise. If M is a Laplace field corresponding to the Laplace measure Λ , the variance process is a gamma process, $\Gamma(\mathbf{s})$, so by the argument above one has that the right hand side of (10) under the finite element approximation and conditionally on the gamma process is $\mathbf{N}(\tilde{\mathbf{m}}, \tilde{\Sigma})$, where the elements of $\tilde{\mathbf{m}}$ and $\tilde{\Sigma}$ are given by

$$\begin{aligned} \tilde{\Sigma}_{ij} &= \mathbf{C} \left(\int_D \varphi_i(\mathbf{s}) \Lambda(d\mathbf{s}), \int_D \varphi_j(\mathbf{s}) \Lambda(d\mathbf{s}) \middle| \Gamma \right) = \int_D \varphi_i(\mathbf{s}) \varphi_j(\mathbf{s}) \Gamma(d\mathbf{s}), \\ \tilde{\mathbf{m}}_i &= \mathbf{E} \left(\int_D \varphi_i(\mathbf{s}) \Lambda(d\mathbf{s}) \middle| \Gamma \right) = \gamma \int_D \varphi_i(\mathbf{s}) d\mathbf{s} + \int_D \varphi_i(\mathbf{s}) \Gamma(d\mathbf{s}). \end{aligned}$$

Given this, the weights \mathbf{w} can be calculated conditionally on the gamma process, $\Gamma(\mathbf{s})$, as

$$\mathbf{w} | \Gamma \sim \mathbf{N} \left(\mathbf{K}_\alpha^{-1} \tilde{\mathbf{m}}, \mathbf{K}_\alpha^{-1} \tilde{\Sigma} \mathbf{K}_\alpha^{-1} \right), \quad (17)$$

where \mathbf{K}_α is defined recursively as in the Gaussian case.

It would seem as one has not gained much by using the conditional representation since the conditional mean and covariances, $\tilde{\mathbf{m}}_i$ and $\tilde{\Sigma}_{ij}$, do not have any simple distributions. One way of approximating them is to approximate the integrals with respect to the Gamma process using the right hand side of (15) with a finite number of terms. However, by using compactly supported linear basis functions, one can simplify things further. Thus, now assume that the basis functions are piecewise linear functions induced by some triangulation of the domain. One can then perform the same Markov approximation as in the Gaussian case. This results in an approximation of the right-hand side of (10) conditionally on the gamma process distributed as $\mathbf{N}(\mathbf{m}, \Sigma)$ with $\mathbf{m} = \gamma\tau\mathbf{a} + \mu\Gamma$, and $\Sigma = \text{diag}(\Gamma)$. Here, the gamma variables $\Gamma_i \sim \Gamma(\tau\mathbf{a}_i, 1)$ are independent and $\mathbf{a}_i = \int \varphi_i(\mathbf{s}) d\mathbf{s}$, and these can be calculated without numerically estimating the integrals with respect to the gamma process.

Bolin and Lindgren (2009) studies how this approximation affects the resulting covariance function of the process in the Gaussian case, and it is shown that the error is small if the approximation is used for piecewise linear basis functions. Although additional studies are needed in the non-Gaussian case, the results are likely similar so that the simplification has no large impact on the approximation. Figures 2-4 show that the approximation is accurate in one and two dimensions as explained in Section 5.

4.2. The solution for general $\alpha > d/2$

If one could approximate the solution to (8) for $\alpha = 1$, the recursive scheme discussed above could be used to represent the solutions for all positive odd α . In the Gaussian case, Lindgren, Rue and Lindström (2011) use a least-squares method where the test functions are chosen as $\psi_i = (\kappa^2 - \Delta)^{\frac{1}{2}}\varphi_i$. The left-hand side of (10) can then be expressed as $\mathbf{K}\mathbf{w}$ and the right-hand side is a mean zero Gaussian variable with covariance matrix \mathbf{K} . This follows from Lemma 2 in Lindgren, Rue and Lindström (2011), which shows that the covariance between element i and element j on the right-hand side can be written as

$$\Sigma_{ij} = \left\langle (\kappa^2 - \Delta)^{\frac{1}{2}}\varphi_i, (\kappa^2 - \Delta)^{\frac{1}{2}}\varphi_j \right\rangle = \left\langle (\kappa^2 - \Delta)\varphi_i, \varphi_j \right\rangle = \mathbf{K}_{ij}.$$

The stochastic weights therefore form a GMRF $\mathbf{w} \sim \mathbf{N}(\mathbf{0}, \mathbf{K}^{-1})$. This argument is unfortunately not applicable in the non-Gaussian case as the covariance between the elements given the gamma process $\Gamma(\mathbf{s})$ is

$$\begin{aligned} \Sigma_{ij} &= \mathbb{C} \left(\int_D (\kappa^2 - \Delta)^{\frac{1}{2}}\varphi_i(\mathbf{s})\Lambda(d\mathbf{s}), \int_D (\kappa^2 - \Delta)^{\frac{1}{2}}\varphi_j(\mathbf{s})\Lambda(d\mathbf{s}) \middle| \Gamma \right) \\ &= \int_D \left((\kappa^2 - \Delta)^{\frac{1}{2}}\varphi_i(\mathbf{s}) \right) \left((\kappa^2 - \Delta)^{\frac{1}{2}}\varphi_j(\mathbf{s}) \right) \Gamma(d\mathbf{s}) \\ &\neq \int_D ((\kappa^2 - \Delta)\varphi_i(\mathbf{s})) \varphi_j(\mathbf{s}) \Gamma(d\mathbf{s}). \end{aligned}$$

We have not been able to find an easy way of evaluating Σ_{ij} in the non-Gaussian case, and it seems as this least-squares procedure is not extendable to the non-Gaussian case. However, if one instead uses $\psi_i = \varphi_i$, the right-hand side of (10) conditionally on the variance process is $N(\mathbf{m}, \Sigma)$, as in the case $\alpha = 2$. With this as a starting point, one can use a finite element matrix transfer technique (FE-MTT) to obtain a discretized approximation of the solution. Simpson (2008) studied such methods for sampling generalized Matérn fields on locally planar Riemannian manifolds, and argued that one could sample the stochastic weights for a general α using the matrix transfer equation $(\mathbf{C}^{-1}\mathbf{K})^{\alpha/2}\mathbf{w} \sim N(0, \mathbf{C}^{-1})$. To simplify the notations in later sections, denote $\mathbf{K}_\alpha = (\mathbf{C}^{-1}\mathbf{K})^{\alpha/2}$ and note that we now have changed the definition of \mathbf{K}_α from the one that was used for even α . The weights \mathbf{w} are then mean zero Gaussian with a precision matrix $\mathbf{Q}_\alpha = \mathbf{K}_\alpha \mathbf{C}^{-1} \mathbf{K}_\alpha$. In the case $\alpha = 2$, this discretization coincides with the approximation described above, but it can be used for any $\alpha > d/2$.

Now in the non-Gaussian case, the results from the case $\alpha = 2$ can be used directly to get a right-hand side that is Gaussian with mean \mathbf{m} and covariance Σ conditionally on the variance process. As in the Gaussian case, this should be multiplied with \mathbf{C}^{-1} to get consistency in the FE-MTT procedure. Hence, in the case of Laplace noise the weights are given by

$$\mathbf{w}|\Gamma \sim N(\mathbf{K}_\alpha^{-1}\mathbf{C}^{-1}\mathbf{m}, \mathbf{K}_\alpha^{-1}\mathbf{C}^{-1}\Sigma\mathbf{C}^{-1}\mathbf{K}_\alpha^{-1}). \quad (18)$$

Again, for the case $\alpha = 2$, this coincides with the procedure described in the section above, and because of this we will from now on use this FE-MTT procedure for all $\alpha > d/2$. Consistency of the FE-MTT procedure follows from similar arguments as in Simpson (2008). These arguments do not provide a rate of convergence as the number of basis functions are increased, and as for the Gaussian case, the rate of convergence and the numerical properties of the approximation are strongly dependent on α .

5. Sampling from the model

Using the finite element representation obtained in the previous section it is easy to generate samples from the SPDE (8). Assume that we want sample the model at locations $\mathbf{s} = (\mathbf{s}_1, \dots, \mathbf{s}_n)$, and let Φ be a matrix with elements $\Phi_{ij} = \varphi_j(\mathbf{s}_i)$. Samples can now be generated using the following three-step algorithm.

Algorithm 5.1. *Sampling the Laplace driven SPDE (8).*

1. Generate two independent random vectors Γ and \mathbf{Z} , where $\Gamma_i \sim \Gamma(\tau\mathbf{a}_i, 1)$ and $\mathbf{Z}_i \sim N(0, 1)$.
2. Let $\Lambda = \gamma\tau\mathbf{a} + \mu\Gamma + \text{diag}(\sqrt{\Gamma})\mathbf{Z}$ and calculate $\mathbf{w} = \mathbf{C}^{-1}\Lambda$.
3. $\mathbf{X} = \Phi\mathbf{K}_\alpha^{-1}\mathbf{w}$ is now a sample of the random field at the locations \mathbf{s} .

The last step could potentially be computationally expensive for large simulations. However, if α is even, one can take advantage of the sparsity of \mathbf{K}_α and solve the equation system $\mathbf{v} = \mathbf{K}_\alpha^{-1}\mathbf{w}$ efficiently without calculating the inverse

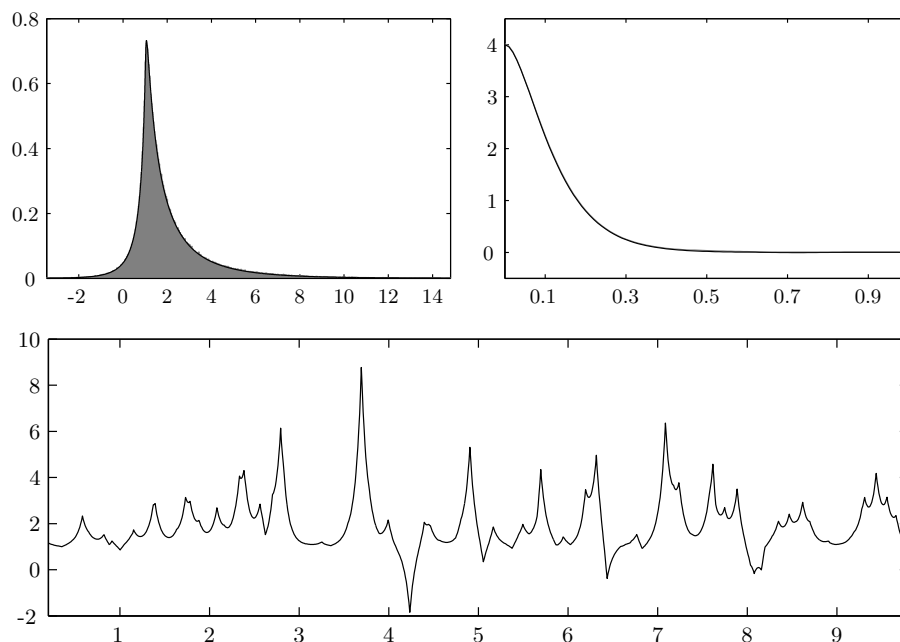


FIGURE 2. The lower panel shows a simulation of the Laplace driven SPDE (8) on \mathbb{R} with parameters $\mu = \gamma = \sigma = 1$, $\tau = 2$, $\kappa = 15$, and $\alpha = 2$. The upper left panel shows a histogram of the samples from 1000 simulations together with the true density. The upper right panel shows the empirical covariance function for the samples (grey curve) together with the true Matérn covariance function (black curve). It is difficult to see the grey curve since the two curves are very similar.

by using Cholesky factorization and back substitution as suggested by [Rue and Held \(2005\)](#). For other $\alpha > d/2$, \mathbf{K}_α is not sparse and the Cholesky method will not improve the computational efficiency. However, as [Simpson \(2008\)](#) shows, one can instead use Krylov subspace methods in the calculations to obtain efficient sampling schemes. The basic problem for general α is to solve the matrix equation $\mathbf{v} = (\mathbf{C}^{-1}\mathbf{K})^{-\frac{\alpha}{2}}\mathbf{w}$, and there are a number of methods with different computational properties that can be used. In this work we use the method by [Hale, Higham and Trefethen \(2008\)](#), which is based on combining contour integrals evaluated by the periodic trapezoid rule with conformal maps involving Jacobi elliptic functions.

In Figure 2, a simulation of a process on \mathbb{R} with parameters $\mu = \gamma = \sigma = 1$, $\tau = 2$, $\kappa = 15$, and $\alpha = 2$ is shown. Since \mathbf{K}_α is sparse in this case, the Cholesky method is used for the simulation. In the upper left panel, a histogram of the samples from 1000 simulations is shown together with the theoretical density, calculated using numerical Fourier inversion of the characteristic function (13). In the upper right panel, the empirical covariance function of the samples is shown together with the theoretical Matérn covariance function. Two more examples of densities and covariance functions for different parameter settings are

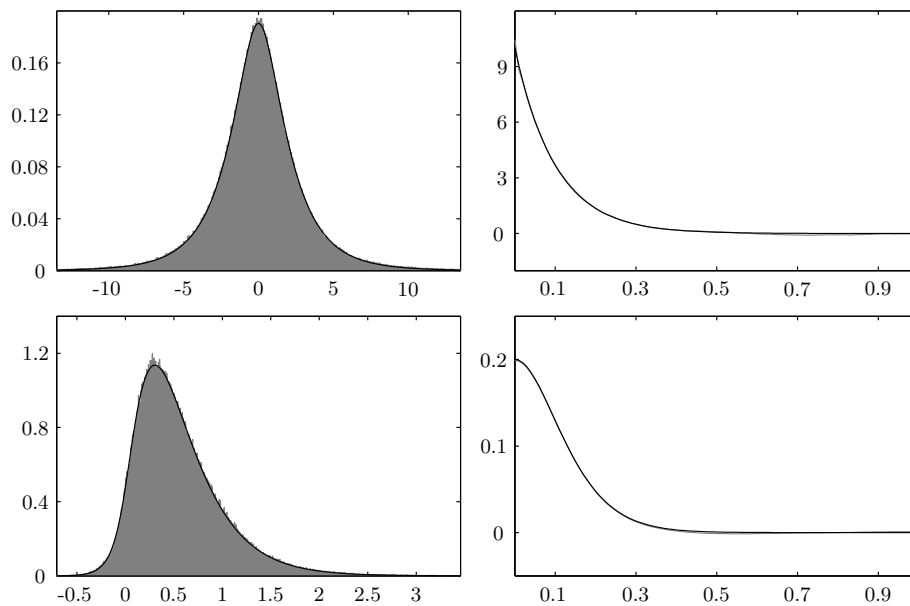


FIGURE 3. Simulation results as in Figure 2 with different parameters. The top row shows a symmetric case with parameters $\mu = \gamma = 0$, $\sigma = 1$, $\kappa = \tau = 10$, and $\alpha = 1$. The bottom row shows an asymmetric case with parameters $\mu = \sigma = 0.1$, $\gamma = 0$, $\kappa = 20$, $\tau = 10$ and $\alpha = 3.5$.

shown in Figure 3. In the upper panels, we have $\alpha = 1$, which results in an exponential covariance function. The other parameters are $\mu = \gamma = 0$, $\sigma = 1$, and $\tau = \kappa = 10$, which results in a symmetric distribution. In the lower panels, we have $\alpha = 3.5$ which results in a smoother field. The other parameters are $\mu = \sigma = 0.1$, $\gamma = 0$, $\tau = 10$, and $\kappa = 20$, which results in an asymmetric distribution. In both cases in Figure 3, the Krylov subspace method is used for the simulations.

In Figure 4 and Figure 5, two simulations of fields on \mathbb{R}^2 are shown together with the corresponding covariance functions, densities, and empirically estimated versions based on 1000 simulations each. As seen in the figures for all five examples, there is a close agreement between the histograms and the true densities, and between the true covariance functions and the empirically estimated covariance functions for all these parameter settings, indicating that the approximation procedure works as intended. A more detailed analysis of the simulation procedure is outside the scope of this article, but it should be noted that the SPDE approximation using piecewise linear basis functions does not provide convergence of higher-order derivatives, and the simulation procedure is therefore not appropriate for applications where such properties are important.

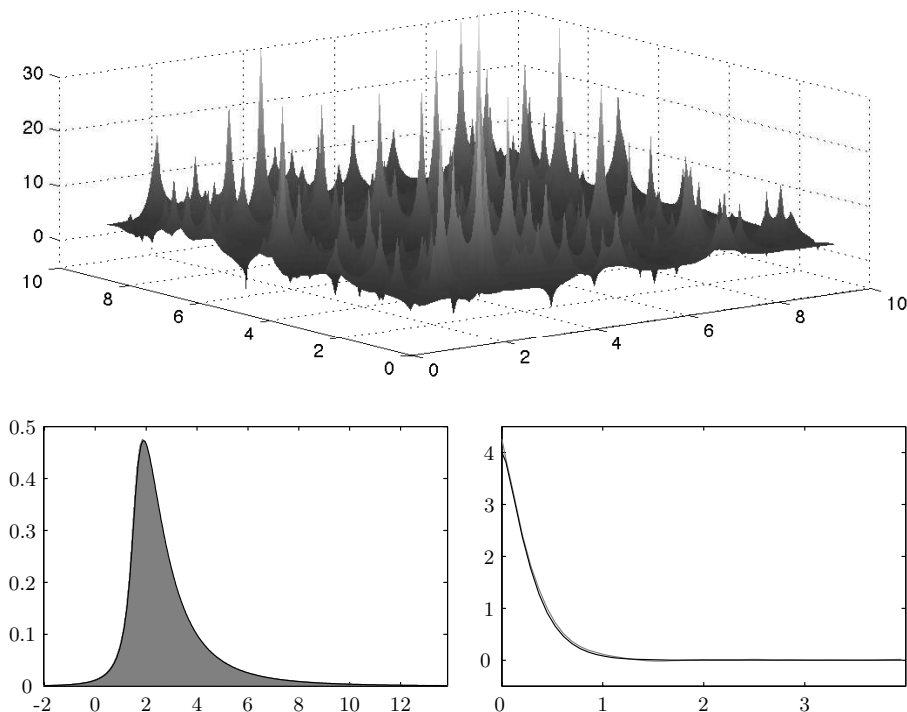


FIGURE 4. A simulation of an asymmetric model (8) in \mathbb{R}^2 where the parameters are $\kappa = 5$, $\sigma = \mu = \gamma = 1$, $\tau = 2$, and $\alpha = 2$. The covariance functions and densities for these fields can be seen in the second row. The empirically estimated versions are based on 1000 simulations.

6. Parameter estimation

Parameter estimation for Laplace moving average models is not easy since there is no closed form expression for the parameter likelihood. Recently, [Podgórski and Wegener \(2011\)](#) derived a method of moments-based estimation procedure for these types of models. In their method, the convolution kernel is first estimated from the spectral density of the data, and given the estimated kernel, the parameters in the Laplace distribution are estimated by fitting the theoretical moments of the Laplace distribution to the sample moments. The method is quite simple although some special care has to be taken to handle the cases when the method of moments equation system does not have a solution, which can happen for certain values of the sample skewness and excess kurtosis.

Using the SPDE formulation, parameter estimation can instead be performed in a likelihood framework. One of the advantages with this is that maximum likelihood parameter estimates always are in the allowed parameter space. Another advantage is that the estimates will account for all relevant information in the data, which might not be the case for method of moment estimates.

To be able to estimate the parameters in a maximum likelihood framework, the problem is interpreted as a missing data problem which facilitates use of the

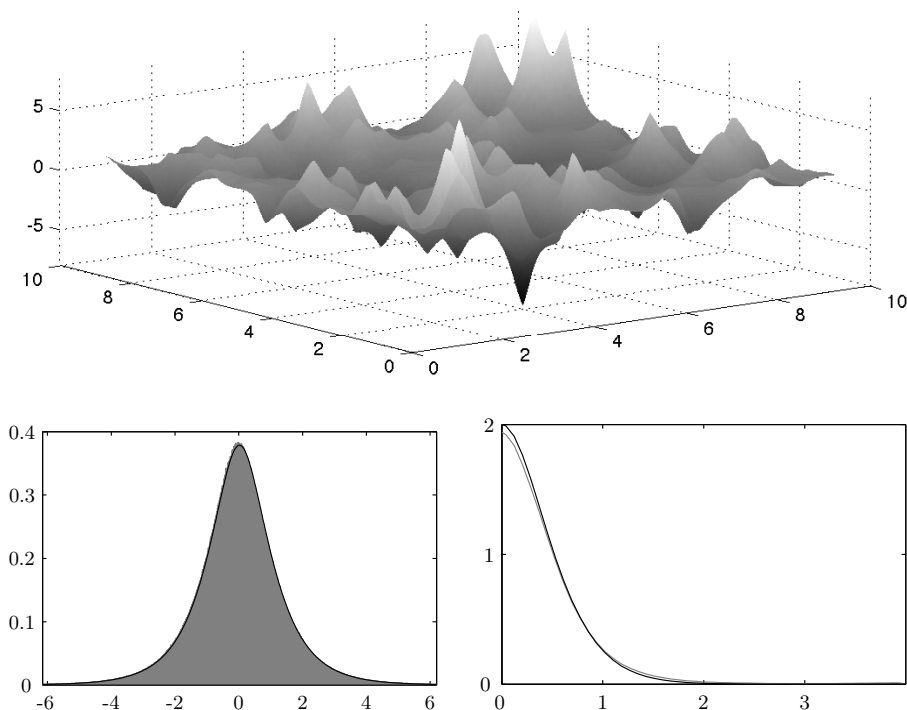


FIGURE 5. A simulation of a symmetric model (8) in \mathbb{R}^2 with parameters $\kappa = 5$, $\sigma = 1$, $\mu = \gamma = 0$, $\tau = 2$, and $\alpha = 4$. The covariance function and density are shown in the second row. The empirically estimated versions are based on 1000 simulations.

Expectation Maximization (EM) algorithm (Dempster, Laird and Rubin, 1977). The proposed EM algorithm is based on the same ideas as the ones in Lange, Little and Taylor (1989) and Protasov (2004) which looked at EM estimation in the case of iid observations of certain Gaussian mixtures. Our main contribution is the extension of these ideas to the random field setting.

Assume we have measurements \mathbf{X} of the process $X(\mathbf{s})$ taken at some locations and that the Hilbert space approximation procedure is used with a basis obtained by triangulating the measurement locations. In this case, the matrix Φ is diagonal and conditioning on the measurements and the parameters is equivalent to conditioning on Λ and the parameters as there is a one-to-one correspondence between the two through $\Lambda = \mathbf{C}\mathbf{K}_\alpha\Phi^{-1}\mathbf{X}$, see Algorithm 5.1. To obtain simpler updating expressions, we first make a change of variables by introducing the parameter $\bar{\gamma} = \gamma\tau$ and estimate this parameter instead of γ . As for Gaussian Matérn models, the shape parameter ν is difficult to estimate accurately and it is therefore assumed to be known throughout this section and no attempt is made at estimating it.

Augmenting the data with the unknown (missing) gamma variables, the augmented likelihood is $L(\boldsymbol{\theta}|\mathbf{X}, \Gamma) = \pi(\mathbf{X}|\Gamma, \boldsymbol{\theta})\pi(\Gamma|\boldsymbol{\theta})$, and the loss-function that

is needed for the EM-procedure is

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(j)}) = \mathbb{E} \left(\log L(\boldsymbol{\theta} | \mathbf{X}, \boldsymbol{\Gamma}) \middle| \mathbf{X}, \boldsymbol{\theta}^{(j)} \right),$$

where $\boldsymbol{\theta}^{(j)}$ is an estimate of $\boldsymbol{\theta} = (\kappa, \sigma, \mu, \bar{\gamma}, \tau)$ at iteration j , and the expectation is taken according to the distribution of $\boldsymbol{\Gamma}$ given \mathbf{X} . We have $\mathbf{X} | \boldsymbol{\Gamma}, \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{m}, \sigma^2 \boldsymbol{\Sigma})$, where $\mathbf{m} = \boldsymbol{\Phi} \mathbf{K}_\alpha^{-1} \mathbf{C}(\bar{\gamma} \mathbf{a} + \mu \boldsymbol{\Gamma})$, $\boldsymbol{\Sigma} = \boldsymbol{\Phi} \mathbf{K}_\alpha^{-1} \mathbf{C} \mathbf{D}_\Gamma \mathbf{C} \mathbf{K}_\alpha^{-1} \boldsymbol{\Phi}$, and \mathbf{D}_Γ is the diagonal matrix with the vector $\boldsymbol{\Gamma}$ on the main diagonal. The second part of the augmented likelihood can be written as $\pi(\boldsymbol{\Gamma} | \boldsymbol{\theta}) = \prod \pi(\Gamma_i | \boldsymbol{\theta})$ since the components in $\boldsymbol{\Gamma}$ are independent gamma variables Γ_i with shape parameters $\tau \mathbf{a}_i$ and scale one, where \mathbf{a}_i are known constants depending on the basis used. The log-likelihood is

$$\begin{aligned} \log L(\boldsymbol{\theta} | \mathbf{X}, \boldsymbol{\Gamma}) &= -n \log(\sigma) + \log(|\mathbf{K}_\alpha|) - \frac{1}{2\sigma^2} (\mathbf{X} - \mathbf{m})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \mathbf{m}) \\ &\quad + \sum_{i=1}^n (\tau \mathbf{a}_i \log \Gamma_i - \log \Gamma(\tau \mathbf{a}_i)) + C, \end{aligned}$$

where the constant C does not depend on the unknown parameters. Thus, using the relation between \mathbf{X} and $\boldsymbol{\Lambda}$, the loss-function is

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(j)}) &= -n \log(\sigma) + \log(|\mathbf{K}_\alpha|) - \frac{1}{2\sigma^2} ((\boldsymbol{\Lambda} - \bar{\gamma} \mathbf{a})^\top \mathbf{D}_{\mathbf{E}(\boldsymbol{\Gamma}^{-1} | \star)} (\boldsymbol{\Lambda} - \bar{\gamma} \mathbf{a}) \\ &\quad + \mu^2 \mathbf{1}^\top \mathbf{E}(\boldsymbol{\Gamma} | \star) + 2\bar{\gamma} \mu \mathbf{a}^\top \mathbf{1} - 2\mu \boldsymbol{\Lambda}^\top \mathbf{1}) \\ &\quad + \sum_{i=1}^n (\tau \mathbf{a}_i \mathbf{E}(\log \Gamma_i | \star) - \log \Gamma(\tau \mathbf{a}_i)) + C, \end{aligned}$$

where $\mathbf{E}(\cdot | \star)$ denotes $\mathbf{E}(\cdot | \boldsymbol{\theta}^{(j)}, \mathbf{X})$. The expectations needed to evaluate the loss-function are $\mathbf{E}(\boldsymbol{\Gamma} | \star)$, $\mathbf{E}(\boldsymbol{\Gamma}^{-1} | \star)$, and $\mathbf{E}(\log \boldsymbol{\Gamma}_i | \star)$. To calculate these, first note that (see [Gradshteyn and Ryzhik, 2000](#), formula 3.472.9)

$$\mathcal{I}(a, b, c) = \int_0^\infty x^{a-1} e^{-\frac{b}{x} - cx} dx = 2 \left(\frac{b}{c} \right)^{\frac{a}{2}} K_a \left(2\sqrt{bc} \right). \quad (19)$$

Using this expression, the expectation $\mathbf{E}(\boldsymbol{\Gamma}_i | \star)$ can be written as

$$\begin{aligned} \mathbf{E}(\boldsymbol{\Gamma}_i | \star) &= \int \boldsymbol{\Gamma}_i \pi(\boldsymbol{\Gamma}_i | \mathbf{X}, \boldsymbol{\theta}) d\boldsymbol{\Gamma}_i = \frac{\int \boldsymbol{\Gamma}_i \pi(\mathbf{X} | \boldsymbol{\Gamma}_i, \boldsymbol{\theta}) \pi(\boldsymbol{\Gamma}_i | \boldsymbol{\theta}) d\boldsymbol{\Gamma}_i}{\pi(\mathbf{X} | \boldsymbol{\theta})} \\ &= \frac{\int \boldsymbol{\Gamma}_i \pi(\mathbf{X} | \boldsymbol{\Gamma}_i, \boldsymbol{\theta}) \pi(\boldsymbol{\Gamma}_i | \boldsymbol{\theta}) d\boldsymbol{\Gamma}_i}{\int \pi(\mathbf{X} | \boldsymbol{\Gamma}_i, \boldsymbol{\theta}) \pi(\boldsymbol{\Gamma}_i | \boldsymbol{\theta}) d\boldsymbol{\Gamma}_i} = \frac{\mathcal{I} \left(\tau \mathbf{a}_i + \frac{1}{2}, \frac{(\boldsymbol{\Lambda}_i - \bar{\gamma} \mathbf{a}_i)^2}{2\sigma^2}, 1 + \frac{\mu^2}{2\sigma^2} \right)}{\mathcal{I} \left(\tau \mathbf{a}_i - \frac{1}{2}, \frac{(\boldsymbol{\Lambda}_i - \bar{\gamma} \mathbf{a}_i)^2}{2\sigma^2}, 1 + \frac{\mu^2}{2\sigma^2} \right)} \\ &= \frac{|\boldsymbol{\Lambda}_i - \bar{\gamma} \mathbf{a}_i|}{\sqrt{2\sigma^2 + \mu^2}} \frac{K_{\tau \mathbf{a}_i + \frac{1}{2}} \left(\sigma^{-2} |\boldsymbol{\Lambda}_i - \bar{\gamma} \mathbf{a}_i| \sqrt{2\sigma^2 + \mu^2} \right)}{K_{\tau \mathbf{a}_i - \frac{1}{2}} \left(\sigma^{-2} |\boldsymbol{\Lambda}_i - \bar{\gamma} \mathbf{a}_i| \sqrt{2\sigma^2 + \mu^2} \right)}. \end{aligned}$$

If the argument in the Bessel functions is very small or very large one might get numerical problems when evaluating this expression depending on how it is implemented. In the case of small arguments, one can use the following approximation to improve the numerical stability

$$K_a(x) \approx \frac{\Gamma(|a|)}{2} \left(\frac{2}{x}\right)^{|a|}, \text{ if } a \neq 0 \text{ and } x \ll \sqrt{|a|+1}.$$

The expectation $\mathbb{E}(\Gamma_i|\star)$ then simplifies to

$$\mathbb{E}(\Gamma_i|\star) \approx \begin{cases} \left(\tau \mathbf{a}_i - \frac{1}{2}\right) \frac{2\sigma^2}{2\sigma^2 + \mu^2}, & \tau > \frac{1}{2\mathbf{a}_i}, \\ \frac{\Gamma(\tau \mathbf{a}_i + \frac{1}{2})}{\Gamma(\frac{1}{2} - \tau \mathbf{a}_i)} \frac{(2\sigma^2)^{2\tau \mathbf{a}_i}}{(2\sigma^2 + \mu^2)^{\frac{2\tau \mathbf{a}_i + 1}{2}}} |\Lambda_i - \bar{\gamma} \mathbf{a}_i|^{1-2\tau \mathbf{a}_i}, & \tau < \frac{1}{2\mathbf{a}_i}. \end{cases}$$

In the case of large arguments, one can instead use the approximation

$$\frac{K_a(x)}{K_{a-1}(x)} \approx 1 + \left(a - \frac{1}{2}\right) \frac{1}{x},$$

which gives the following approximation for $\mathbb{E}(\Gamma_i|\star)$

$$\mathbb{E}(\Gamma_i|\star) \approx \frac{|\Lambda_i - \bar{\gamma} \mathbf{a}_i|}{\sqrt{2\sigma^2 + \mu^2}} + \frac{\tau \mathbf{a}_i \sigma^2}{2\sigma^2 + \mu^2}.$$

The expectation $\mathbb{E}(\Gamma_i^{-1}|\star)$ is calculated similarly using (19) and can be written as

$$\mathbb{E}(\Gamma_i^{-1}|\star) = \frac{\sqrt{2\sigma^2 + \mu^2}}{|\Lambda_i - \bar{\gamma} \mathbf{a}_i|} \frac{K_{\tau \mathbf{a}_i - \frac{3}{2}} \left(\sigma^{-2} |\Lambda_i - \bar{\gamma} \mathbf{a}_i| \sqrt{2\sigma^2 + \mu^2} \right)}{K_{\tau \mathbf{a}_i - \frac{1}{2}} \left(\sigma^{-2} |\Lambda_i - \bar{\gamma} \mathbf{a}_i| \sqrt{2\sigma^2 + \mu^2} \right)}. \quad (20)$$

Evaluating modified Bessel functions numerically is computationally expensive and should therefore be avoided as much as possible when implementing the estimation procedure. To that end, one can express $K_{\tau \mathbf{a}_i - \frac{3}{2}}(\cdot)$ using the following recurrence relationship for modified Bessel functions

$$K_a(x) = K_{a+2}(x) - \frac{2(a+1)}{x} K_{a+1}(x),$$

giving the following expression for $\mathbb{E}(\Gamma_i^{-1}|\star)$ in terms of $\mathbb{E}(\Gamma_i|\star)$

$$\mathbb{E}(\Gamma_i^{-1}|\star) = \frac{(\mu^2 + 2\sigma^2)\mathbb{E}(\Gamma_i|\star) - \sigma^2(2\tau \mathbf{a}_i - 1)}{(\Lambda_i - \bar{\gamma} \mathbf{a}_i)^2}.$$

Using this expression instead of (20), one only has to evaluate two modified Bessel functions instead of three.

Finally, the expectation $\mathbb{E}(\log(\Gamma_i)|\star)$ is similarly written as

$$\mathbb{E}(\log(\Gamma_i)|\star) = \frac{\int \log(\Gamma_i) \pi(\mathbf{X}|\Gamma_i, \boldsymbol{\theta}) \pi(\Gamma_i|\boldsymbol{\theta}) d\Gamma_i}{\pi(\mathbf{X}|\boldsymbol{\theta})}.$$

The denominator is the same as in the previous expectations, while calculating the nominator requires evaluating an integral on the form

$$\mathcal{I}_{\log}(a, b, c) = \int_0^\infty \log(x) x^{a-1} \exp\left(-\frac{b}{x} - cx\right) dx. \quad (21)$$

To calculate this integral, we differentiate (19) with respect to a and obtain

$$\begin{aligned} \mathcal{I}_{\log}(a, b, c) &= \frac{\partial}{\partial a} \int_0^\infty x^{a-1} e^{-\frac{b}{x} - cx} dx = \frac{\partial}{\partial a} \left(2 \left(\frac{b}{c} \right)^{\frac{a}{2}} K_a(2\sqrt{bc}) \right) \\ &= 2 \left(\frac{b}{c} \right)^{\frac{a}{2}} \left(\log\left(\frac{b}{c}\right) K_a(2\sqrt{bc}) + \frac{\partial}{\partial a} K_a(2\sqrt{bc}) \right). \end{aligned}$$

The derivative of $K_a(2\sqrt{bc})$ with respect to a can be expressed using infinite sums of gamma- and polygamma functions; however, in this case it is easier to numerically approximate the derivative using for example forward differences:

$$\frac{\partial}{\partial a} K_a(2\sqrt{bc}) \approx \frac{K_{a+\epsilon}(2\sqrt{bc}) - K_a(2\sqrt{bc})}{\epsilon}.$$

Using this expression, we approximate $E(\log(\Gamma_i)|\star)$ as

$$\begin{aligned} E(\log(\Gamma_i)|\star) &= \frac{\mathcal{I}_{\log}\left(\tau \mathbf{a}_i - \frac{1}{2}, \frac{(\mathbf{\Lambda}_i - \bar{\gamma} \mathbf{a}_i)^2}{2\sigma^2}, 1 + \frac{\mu^2}{2\sigma^2}\right)}{\mathcal{I}\left(\tau \mathbf{a}_i - \frac{1}{2}, \frac{(\mathbf{\Lambda}_i - \bar{\gamma} \mathbf{a}_i)^2}{2\sigma^2}, 1 + \frac{\mu^2}{2\sigma^2}\right)} \\ &\approx \log\left(\frac{|\mathbf{\Lambda}_i - \bar{\gamma} \mathbf{a}_i|}{\sqrt{\mu^2 + 2\sigma^2}}\right) - \frac{1}{\epsilon} \\ &\quad + \frac{1}{\epsilon} \frac{K_{\tau \mathbf{a}_i - \frac{1}{2} + \epsilon}\left(\sigma^{-2} |\mathbf{\Lambda}_i - \bar{\gamma} \mathbf{a}_i| \sqrt{2\sigma^2 + \mu^2}\right)}{K_{\tau \mathbf{a}_i - \frac{1}{2}}\left(\sigma^{-2} |\mathbf{\Lambda}_i - \bar{\gamma} \mathbf{a}_i| \sqrt{2\sigma^2 + \mu^2}\right)}. \end{aligned}$$

To obtain the updating equations for the parameters, the loss-function should be maximized with respect to each parameter, for example by differentiating it with respect to the parameters and setting the derivatives equal to zero. Since the system of equations obtained from this procedure is not analytically solvable, one would have to iterate numerically in each step to obtain the parameter updates if the EM algorithm is used without modifications. A better alternative is to use an Expectation Conditional Maximization (ECM) algorithm (Meng and Rubin, 1993) where the M-step is divided into two conditional maximization steps. In the first step, the parameters of the Laplace noise is updated conditionally on the current value of κ , and in the second step κ is updated conditionally on the other parameters. Differentiating the loss-function with respect to μ , $\bar{\gamma}$, and σ and setting the derivatives equal to zero yields the following

updating rules

$$\begin{aligned}\mu^{(j+1)} &= \frac{(\mathbf{\Lambda}^\top \mathbf{1})(\mathbf{a}^\top \mathbf{D}_{\mathbf{E}(\Gamma^{-1}|\star)} \mathbf{a}) - (\mathbf{a}^\top \mathbf{1})(\mathbf{\Lambda}^\top \mathbf{D}_{\mathbf{E}(\Gamma^{-1}|\star)} \mathbf{a})}{(\mathbf{1}^\top \mathbf{E}(\Gamma|\star))(\mathbf{a}^\top \mathbf{D}_{\mathbf{E}(\Gamma^{-1}|\star)} \mathbf{a}) - (\mathbf{1}^\top \mathbf{a})^2}, \\ \bar{\gamma}^{(j+1)} &= \frac{(\mathbf{1}^\top \mathbf{E}(\Gamma|\star))(\mathbf{\Lambda}^\top \mathbf{D}_{\mathbf{E}(\Gamma^{-1}|\star)} \mathbf{a}) - (\mathbf{\Lambda}^\top \mathbf{1})(\mathbf{a}^\top \mathbf{1})}{(\mathbf{1}^\top \mathbf{E}(\Gamma|\star))(\mathbf{a}^\top \mathbf{D}_{\mathbf{E}(\Gamma^{-1}|\star)} \mathbf{a}) - (\mathbf{1}^\top \mathbf{a})^2}, \\ \sigma^{(j+1)} &= \frac{1}{\sqrt{n}} \left(\mathbf{\Lambda}^\top \mathbf{D}_{\mathbf{E}(\Gamma^{-1}|\star)} \mathbf{\Lambda} + 2 \frac{(\mathbf{\Lambda}^\top \mathbf{D}_{\mathbf{E}(\Gamma^{-1}|\star)} \mathbf{a})(\mathbf{\Lambda}^\top \mathbf{1})(\mathbf{1}^\top \mathbf{a})}{(\mathbf{1}^\top \mathbf{E}(\Gamma|\star))(\mathbf{a}^\top \mathbf{D}_{\mathbf{E}(\Gamma^{-1}|\star)} \mathbf{a}) - (\mathbf{1}^\top \mathbf{a})^2} \right. \\ &\quad \left. - \frac{(\mathbf{\Lambda}^\top \mathbf{D}_{\mathbf{E}(\Gamma^{-1}|\star)} \mathbf{a})^2 (\mathbf{1}^\top \mathbf{E}(\Gamma|\star)) + (\mathbf{a}^\top \mathbf{D}_{\mathbf{E}(\Gamma^{-1}|\star)} \mathbf{a})(\mathbf{\Lambda}^\top \mathbf{1})^2}{(\mathbf{1}^\top \mathbf{E}(\Gamma|\star))(\mathbf{a}^\top \mathbf{D}_{\mathbf{E}(\Gamma^{-1}|\star)} \mathbf{a}) - (\mathbf{1}^\top \mathbf{a})^2} \right)^{\frac{1}{2}}.\end{aligned}$$

In general, there is no closed form expression for the conditional updating equation for τ , so the following equation is maximized numerically to obtain $\tau^{(j+1)}$

$$\mathcal{Q}_\tau = \sum_{i=1}^n (\tau \mathbf{a}_i \mathbf{E}(\log \Gamma_i|\star) - \log \Gamma(\tau \mathbf{a}_i)).$$

In the special case when all \mathbf{a}_i are equal to some value a , which for example is the case if a triangulation induced by a regular lattice is used in the Hilbert space approximation, the solution can be written as

$$\tau^{(j+1)} = \frac{1}{a} \psi^{-1} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{E}(\log \Gamma_i|\star) \right),$$

where $\psi^{-1}(\cdot)$ is the inverse of the digamma function. Finally κ is updated conditionally on the other parameters. There is no closed form expression for the updating equation for κ either, so the following expression is maximized numerically with respect to κ ,

$$\begin{aligned}\mathcal{Q}_\kappa &= \log(|\mathbf{K}_\alpha|) - \frac{1}{2(\sigma^{(j+1)})^2} \left(\mathbf{\Lambda}^\top \mathbf{D}_{\mathbf{E}(\Gamma^{-1}|\star)} \mathbf{\Lambda} \right. \\ &\quad \left. - 2\bar{\gamma}^{(j+1)} \mathbf{\Lambda}^\top \mathbf{D}_{\mathbf{E}(\Gamma^{-1}|\star)} \mathbf{a} - 2\mu^{(j+1)} \mathbf{\Lambda}^\top \mathbf{1} \right).\end{aligned}$$

By the construction of \mathbf{K}_α , its log-determinant can be written as

$$\log(|\mathbf{K}_\alpha|) = \frac{\alpha}{2} \log |\mathbf{C}^{-1} \mathbf{G} + \kappa^2 \mathbf{I}| = \frac{\alpha}{2} \sum_{i=1}^n \log(\lambda_i + \kappa^2),$$

where λ_i denotes the i th eigenvalue of $\mathbf{C}^{-1} \mathbf{G}$. If the size of \mathbf{K}_α is small, these eigenvalues can be pre-calculated as they do not depend on the parameters. For larger problems it is most efficient to calculate the log-determinant in each iteration using a sparse Cholesky factorization of $\mathbf{K} = \mathbf{G} + \kappa^2 \mathbf{C}$.

As shown by [Meng and Rubin \(1993\)](#), the ECM algorithm has the same convergence properties as the ordinary EM algorithm. The likelihood is increasing for each iteration and the convergence is linear. Hence, we do not lose any rate of convergence by using the ECM algorithm instead of the EM algorithm.

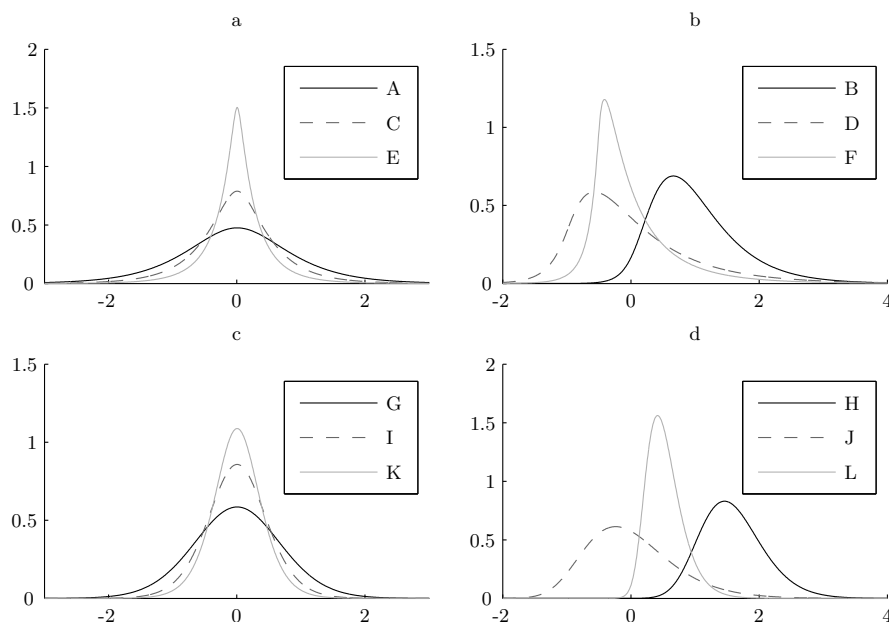


FIGURE 6. Marginal distributions for the twelve test cases in the simulation study. In Panel a and Panel b, the approximate covariance range is 3.5 and in Panel c and d, the range is 35.

7. A simulation study

In this section, a simulation study is performed to test the accuracy of the parameter estimation algorithm presented above. The algorithm is tested for twelve different parameter settings corresponding to marginal distributions shown in Figure 6 for processes in one dimension with $\alpha = 2$. For Matérn covariance functions, one sometimes defines the approximate range as $r = \sqrt{8\nu\kappa^{-1}}$, which is the value where the correlation is approximately 0.1. For the first six test cases, we have $\kappa = 1$ which corresponds to an approximate range of 3.5, and for the last six cases we have $\kappa = 0.1$ which corresponds to an approximate range of 35. For each value of κ , three symmetric distributions and three asymmetric distributions are used. In Figure 6, the distributions for the short range are shown in the two upper panels, and the distributions for the long range are shown in the two bottom panels.

For each set of parameters, 500 data sets are simulated using Algorithm 5.1, where each data set contains 1000 equally spaced observations on $[1, 1000]$. The basis used in the Hilbert space approximations consists of 1000 piecewise linear basis functions centered at $1, 2, \dots, 1000$. For each data set, the starting value for κ is set to $\sqrt{8\nu\hat{r}^{-1}}$, where \hat{r} is the approximate range for the empirical covariance function for the data set. To obtain good starting values for the other parameters, an initial run of the EM estimator is made with κ fixed to the

	κ	τ	σ	μ	γ
A	1 (0.95 1.00 1.06)	2 (1.63 2.03 3.02)	1 (0.78 0.98 1.13)	0 (-0.07 -0.01 0.06)	0 (-0.06 0.00 0.07)
B	1 (0.96 1.00 1.05)	2 (1.68 1.99 2.41)	$\frac{1}{2}$ (0.42 0.50 0.57)	$\frac{1}{2}$ (0.43 0.50 0.57)	0 (-0.06 0.00 0.07)
C	1 (0.96 1.00 1.05)	1 (0.85 0.99 1.21)	1 (0.87 1.00 1.11)	0 (-0.07 0.00 0.05)	0 (-0.04 0.00 0.05)
D	1 (0.96 1.00 1.04)	1 (0.90 1.00 1.14)	1 (0.90 1.00 1.10)	1 (0.87 1.00 1.13)	-1 (-1.11 -1.00 -0.89)
E	1 (0.97 1.00 1.02)	$\frac{1}{2}$ (0.45 0.49 0.54)	1 (0.93 1.00 1.08)	0 (-0.06 0.00 0.06)	0 (-0.01 0.00 0.01)
F	1 (0.98 1.00 1.01)	$\frac{1}{2}$ (0.46 0.50 0.54)	1 (0.91 1.00 1.08)	1 (0.89 1.00 1.11)	-1 (-1.09 -1.01 -0.92)
G	$\frac{1}{10}$ (0.09 0.10 0.11)	1 (0.86 1.00 1.24)	1 (0.87 1.00 1.11)	0 (-0.06 0.00 0.06)	0 (-0.04 0.00 0.04)
H	$\frac{1}{10}$ (0.09 0.10 0.11)	1 (0.89 0.99 1.13)	$\frac{1}{2}$ (0.45 0.50 0.54)	$\frac{1}{2}$ (0.44 0.50 0.56)	0 (-0.05 0.01 0.13)
I	$\frac{1}{10}$ (0.10 0.10 0.10)	$\frac{1}{2}$ (0.45 0.49 0.54)	1 (0.91 1.01 1.09)	0 (-0.06 0.00 0.07)	0 (-0.01 0.00 0.01)
J	$\frac{1}{10}$ (0.10 0.10 0.10)	$\frac{1}{2}$ (0.46 0.50 0.54)	1 (0.91 0.99 1.08)	1 (0.90 1.01 1.13)	-1 (-1.09 -1.01 -0.93)
K	$\frac{1}{10}$ (0.10 0.10 0.10)	$\frac{1}{3}$ (0.31 0.33 0.36)	1 (0.91 0.99 1.06)	0 (-0.07 0.00 0.07)	0 (-0.00 0.00 0.00)
L	$\frac{1}{10}$ (0.09 0.10 0.12)	$\frac{1}{3}$ (0.33 0.36 0.43)	$\frac{1}{2}$ (0.42 0.47 0.51)	$\frac{1}{2}$ (0.39 0.50 0.52)	0 (-0.05 0.00 0.13)

TABLE 1

Parameter settings for the twelve cases shown in Figure 6 the estimation procedure is tested for. In the parentheses, the 10% 50%, and 90% percentiles of 500 Monte Carlo samples are shown. Note that most estimates seem to be unbiased, perhaps with the exception of the estimates of τ in case L.

starting value and where the starting values for μ and γ are drawn independently from a $N(0, 1)$ distribution, and the starting values for σ and τ^{-1} are drawn from a $\chi^2(1)$ -distribution. After 100 steps, this initial run is ended, and the estimates are used as starting values for the full EM-estimator.

In Table 1, the 10% 50%, and 90% percentiles of 500 Monte Carlo samples are shown for each parameter setting, together with the true values of the parameters. One can note that all estimates are more or less unbiased and have fairly small variances, indicating that the estimation procedure works as intended. The only case where the estimator seems to have a bias is in case L, where most of the estimated values of τ are above the true value. The cause of this bias is probably that the estimation procedure is not very stable for small values of τ because some of the expectations $E(\Gamma_i^{-1}|\star)$ can be infinite in this case. More precisely, for $\tau < 3/(2 \min(\mathbf{a}_1, \dots, \mathbf{a}_n))$, the likelihood is unbounded for any $\bar{\gamma} = \Lambda_i/\mathbf{a}_i$ and the ML procedure thus has to be modified. To improve the stability of the algorithm, the expectations $E(\Gamma_i^{-1}|\star)$ are truncated to 1000 in the first iteration, and for each iteration this bound is made larger so that it has little to no effect after a few hundred iterations of the algorithm. This greatly improves the stability for $\tau < 3/(2 \min(\mathbf{a}_1, \dots, \mathbf{a}_n))$, but it is left for future research to justify this modified maximum likelihood procedure theoretically, to derive large sample properties of the estimator, and to investigate other improvements for the case of small values of τ .

It should finally be noted that the parameters are estimated assuming the same finite element approximation as is used for simulating the data. Estimating the model parameters using a different numbers of basis functions in the approximation can possibly give biased estimates, as the parameters are estimated to maximize the likelihood for the approximate model instead of the exact SPDE. The size of this bias depends on the specific parameters of the model, and especially on the true covariance range in relation to the spacing of the

basis functions, as discussed in [Bolin and Lindgren \(2011\)](#) in the case of Gaussian models. It is, however, outside the scope of this work to investigate this issue further here.

8. Discussion and extensions

We have showed how the SPDE approach by [Lindgren, Rue and Lindström \(2011\)](#) can be extended to the case of Laplace noise and how this can be used to obtain an efficient estimation procedure as well as an accurate estimation technique for the Laplace moving average models. This is indented as a demonstration that the methods in [Lindgren, Rue and Lindström \(2011\)](#) are applicable to more general situations than the ordinary Gaussian models. There are also a number of extensions that can be made to this work which are discussed below.

First of all, the Hilbert space approximation technique in [Section 4](#) was derived using theory for Lévy processes of type G, and although we only used this for the case of Laplace noise, the methods work equally well for this larger class of models. All that is changed are the distributions of the integrals conditionally on the variance process. These techniques are also applicable to the case when more general SPDEs are used, one could for example use the nested SPDEs by [Bolin and Lindgren \(2011\)](#) to achieve more general covariance structures without any additional work needed, or one could include drift terms in the operator on the left-hand side to mimic the effects of asymmetric kernels in the Laplace moving average models. The methods are in fact not restricted to \mathbb{R}^d or stationary SPDEs, but can be extended to non-stationary SPDEs on general Riemann manifolds.

Secondly, the estimation procedure in [Section 6](#) assumed that one basis function was used for each observation of the process. The reason being that this gives us a one-to-one correspondence between the observations and the Laplace variables $\mathbf{\Lambda}$ which simplified the estimation procedure. For practical applications this is not ideal as one would like to be able to choose the basis independently of the measurement locations, and it would also be useful if one could assume that the measurements are taken under measurement noise. If the estimation procedure could be extended to handle these cases, the practical usefulness of these models would greatly improve.

As mentioned in [Section 7](#), the estimation procedure is sensitive to the value of τ . Too large values will result in a model which is very similar to a standard Gaussian model, and it might be difficult to accurately estimate the parameters in this case without a very large data set. This is not a big problem as if the data is Gaussian, one should not use these models but a standard Gaussian model. The estimation procedure is also unstable for small values of τ , and modifications to further improve the stability in this case are currently being investigated.

Acknowledgements

The author is grateful to Krzysztof Podgórski for many helpful comments and discussions regarding the theoretical aspects of this work, to Daniel Simpson for providing some of his code for the Krylov subspace methods used in Section 5, and to Jonas Wallin for numerous discussions regarding the parameter estimation problem and for suggesting the truncation of the expectations mentioned at the end of Section 7.

References

- ÅBERG, S., PODGÓRSKI, K. and RYCHLIK, I. (2009). Fatigue damage assessment for a spectral model of non-Gaussian random loads. **24** 608-617.
- ÅBERG, S. and PODGÓRSKI, K. (2011). A class of non-Gaussian second order random fields. *Extremes* **14** 187-222.
- ADAMS, R. A. (1975). *Sobolev Spaces*. Academic Press.
- BOGSJÖ, K., PODGÓRSKI, K. and RYCHLIK, I. (2012). Models for road surface roughness. *Vehicle System Dynamics* **50** 725-747.
- BOLIN, D. and LINDGREN, F. (2009). Wavelet Markov approximations as efficient alternatives to tapering and convolution fields (submitted). **2009:13**.
- BOLIN, D. and LINDGREN, F. (2011). Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *Ann. Appl. Statist.* **5** 523-550.
- DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1-38.
- GRADSHTEYN, I. S. and RYZHIK, I. M. (2000). *Table of integrals, series and products*, 6 ed. Elsevier Inc.
- HALE, N., HIGHAM, N. J. and TREFETHEN, L. N. (2008). Computing A^α , $\log(A)$ and Related Matrix Functions by Contour Integrals. **46** 2505-2523.
- HIGDON, D. (2001). Space and Space-time modeling using process convolutions Technical Report.
- LANGE, K., LITTLE, R. and TAYLOR, J. (1989). Robust statistical modeling using the t distribution. *J. Amer. Statist. Assoc.* **84** 881-896.
- LINDGREN, F., RUE, H. and LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). *J. Roy. Statist. Soc. Ser. B* **73** 423-498.
- MATÉRN, B. (1960). Spatial variation. *Meddelanden från statens skogsforskningsinstitut* **49**.
- MENG, X.-L. and RUBIN, D. B. (1993). Maximum Likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80** 267-78.
- PODGÓRSKI, K. and WEGENER, J. (2011). Estimation for stochastic models driven by Laplace motion. **40** 3281-3302.

- PROTASSOV, R. (2004). EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ . *Statist. and Comput.* **14** 67–77.
- RØISLIEN, J. and OMRE, H. (2006). T-distributed random fields: A parametric model for heavy-tailed well-log data. *Math. Geol.* **38** 821–849.
- ROSÍŃSKI, J. (1991). On a class of infinitely divisible processes represented as mixtures of Gaussian processes. In *Stable Processes and Related Topics. Progress in Probability* **25** 27–41. Birkhauser, Boston.
- RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields; Theory and Applications. Monographs on Statistics and Applied Probability* **104**. Chapman & Hall/CRC.
- SAMKO, S. G., KILBAS, A. A. and MARICEV, O. I. (1992). *Fractional integrals and derivatives: theory and applications*. Gordon and Breach Science Publishers, Yveron.
- SIMPSON, D. P. (2008). Krylov subspace methods for approximating functions of symmetric positive definite matrices with applications to applied statistics and anomalous diffusion. *PhD thesis, Queensland University of Technology*.
- SIMPSON, D., LINDGREN, F. and RUE, H. (2010). In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Preprint, statistics, Trondheim, Norway* **16/2010**.
- WALSH, J. (1986). An introduction to stochastic partial differential equations. In *École d'Été de Probabilités de Saint Flour XIV - 1984. Lecture Notes in Mathematics* **1180** 3 265–439. Springer Berlin / Heidelberg.
- WHITTLE, P. (1963). Stochastic processes in several dimensions. *Bull. Internat. Statist. Inst.* **40** 974–994.
- WIKTORSSON, M. (2002). Simulation of stochastic integrals with respect to Lévy processes of type G. **101** 113–125.